

Algoritmos supervisados de machine learning para determinar la ubicación de dispositivos wifi

Jahaira Macias

17 de octubre de 2023

Resumen

El presente artículo tiene como objetivo elegir, bajo un determinado escenario, el mejor algoritmo supervisado de machine learning para localizar un terminal que soporte wifi. Se usa un dataset que cuenta con 2000 registros de Received Signal Strength Indicator (RSSI), obtenidos de 7 puntos de acceso (AP), los cuales se cargan en 8 algoritmos supervisados de machine learning. Luego se elige el algoritmo que realiza la predicción más precisa, incluso cuando se cuenta con un menor número de AP.

1. Introducción

Determinar la ubicación de objetos o dispositivos en interiores y exteriores utilizando redes inalámbricas, específicamente tecnología wifi. Se mencionan diversas aplicaciones de esta tecnología, como el monitoreo de productos y personas en situaciones de emergencia.

El artículo se centra en el uso del parámetro RSSI (Received Signal Strength Indicator) como una forma de predecir la ubicación de dispositivos. Xun et al. (2020) [6] indica que, aunque el parámetro CSI (Channel State Information) también es relevante, su uso requiere hardware específico y modificaciones en el firmware, por lo que se prefiere el RSSI debido a su disponibilidad en dispositivos como smartphones.

Para este trabajo se utilizó una herramienta de código abierto KNIME para explorar el uso de algoritmos de machine learning en la determinación de la ubicación de un smartphone conectado a una red wifi. La metodología incluye la comparación de ocho algoritmos supervisados, desde árboles de decisión hasta redes neuronales, utilizando KNIME.

En específico, se usarán ocho algoritmos, incluyendo:

- decisión tree (DT)
- naive Bayes (NB)
- k-nearest neighbors (k-NN)

- random forest (RF)
- support vector machine (SVM)
- gradient boosted (GB)

así como algoritmos basados en redes neuronales:

- RProp MLP (multilayer perceptron)
- PNN (probabilistic neural network)

Con el fin de profundizar en los algoritmos utilizados, se sugiere revisar KNIME (2020)

2. Estado Del Arte

Se menciona que en varios artículos se han enfocado en determinar la ubicación de objetos compatibles con wifi utilizando información de puntos de acceso (AP) y algoritmos de machine learning. Se hace referencia a un estudio de Bellavista-Parent et al. (2021) [1] que revisó 64 artículos relacionados con el posicionamiento indoor a lo largo de 2019 y 2021. Además, se llevó a cabo una nueva búsqueda en Scopus y Web of Science, donde se encontraron 18 artículos relacionados con el posicionamiento mediante wifi y algoritmos de machine learning.

Se seleccionaron 10 de estos artículos para su inclusión en el estudio debido a su relevancia. En algunos de los hallazgos de estos artículos, Por ejemplo, Sabanci et al. (2018) [2] compararon varios algoritmos y encontraron que k-NN ofrecía el mejor rendimiento en la determinación de ubicaciones en cuatro habitaciones. Tahat et al. (2021) [4] utilizaron redes neuronales y encontraron que estas presentaban buenos resultados, aunque requerían un mayor tiempo de entrenamiento. Además, se hace referencia a un nuevo enfoque llamado "localización indoor de alta adaptabilidad" (HAIL) desarrollado por Xue et al. (2020) [5], que utiliza una red neuronal de retropropagación (BPNN) y sugiere la necesidad de investigar nuevos algoritmos para mejorar la precisión.

Finalmente, se menciona un estudio de Singh et al. (2021) [3] que revisó artículos publicados entre 2007 y 2020 relacionados con el uso de algoritmos de machine learning para determinar ubicaciones indoor y planteó preocupaciones sobre la privacidad en la recolección de información en un mundo cada vez más conectado.

3. Metodología

Para hacer la comparación de cada uno de los algoritmos usados en la investigación se uso un dataset llamado Wireless Indoor Localization, El dataset contiene 2000 mediciones de intensidad de señal de recepción en un smartphone, distribuidas en 4 habitaciones con 500 muestras cada una. Para ello, el

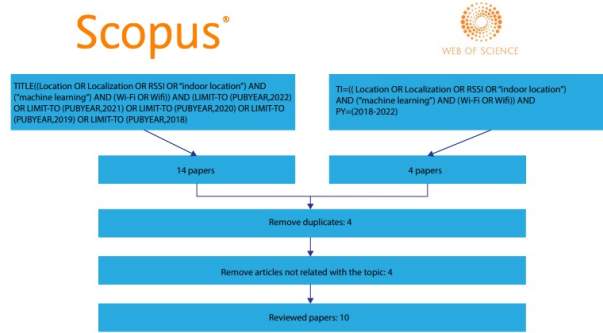


Figura 1: Artículos revisados en Scopus y Web of Science.

smartphone se desplazó por 4 habitaciones, recogiendo las mediciones del RSSI de 7 AP. Cabe señalar que no se tiene mayor detalle del dataset, por ejemplo, banda de frecuencia, modelo del terminal móvil, canales utilizados, entre otros.

AP_01	AP_02	AP_03	AP_04	AP_05	AP_06	AP_07	Room
-64	-56	-61	-66	-71	-82	-81	Room_1
-68	-57	-61	-65	-71	-85	-85	Room_1
-63	-60	-60	-67	-76	-85	-84	Room_1
-34	-53	-52	-19	-65	-75	-69	Room_2

Figura 2: Aquí podemos ver un extracto del dataset utilizado en esta investigación.

El dataset, que contiene mediciones de diversos AP, se dividirá en tres tipos. En primer lugar, se usará el dataset original (con 7 AP), luego con 6 AP y, finalmente, solo considerando 5 AP. Esto se realiza con el fin de validar el comportamiento del algoritmo ante fallas de un determinado AP. Se dividió un dataset de mediciones de puntos de acceso (AP) en tres conjuntos: uno con 7 AP originales, otro con 6 AP y otro con 5 AP, para evaluar el comportamiento del algoritmo ante la falta de un AP como se ve en la Figura 3. Así mismo se normalizó la señal RSSI en un rango de 0 a 100 y se dividió el dataset en 70 % para entrenamiento y 30 % para evaluar la precisión del algoritmo. Se seleccionó el algoritmo con el menor error (mayor precisión) para su uso.

En la siguiente figura nos encontramos con el modelo que parte desde la formación de tres grupos de mediciones (datos de ingreso del modelo) hasta llegar al algoritmo que presente el menor error de predicción (datos de salida del modelo).

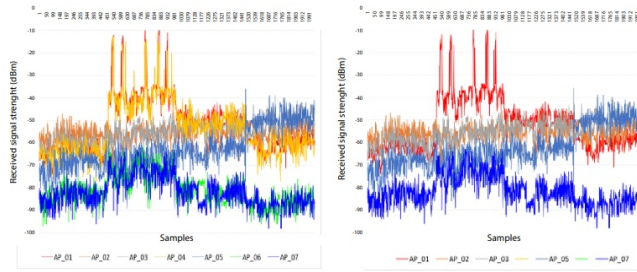


Figura 3: Niveles de intensidad de señal recibidos en el smartphone usando datos de 7 y 5 AP.

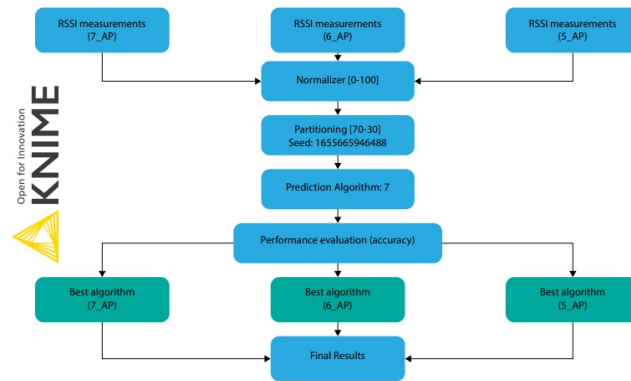


Figura 4: Modelo utilizado para evaluar la precisión de los algoritmos.

4. Implementación de los algoritmos de machine learning en KNIME

El flujo de trabajo en KNIME se compone de tres datasets que se cargan de forma independiente. Los datos ingresan a través del módulo CSV Reader y los resultados se obtienen en el bloque "Scorer". Antes de la partición, se normaliza el dataset en un rango de 0 a 100. Luego, se divide el conjunto de datos en una proporción de 70/30, asegurando que cada habitación tenga la misma cantidad de muestras. Se utiliza una semilla específica (1655665946488) para garantizar la reproducibilidad de los resultados.

El flujo de trabajo involucra la aplicación de ocho algoritmos no supervisados para resolver un problema de clasificación que busca detectar la habitación en la que se encuentra un terminal móvil. Estos algoritmos incluyen Decision Trees (DT), Naïve Bayes (NB), k-Nearest Neighbors (k-NN), Random Forest (RF), Support Vector Machines (SVM), Gradient Boosting (GB) y dos algoritmos basados en redes neuronales artificiales (ANN), específicamente RProp MLP y

PNN.

Cada algoritmo de machine learning se incorpora al flujo de trabajo como un bloque independiente debido a la facilidad de uso de KNIME. Luego, se evalúa la precisión de cada algoritmo en el módulo "Scorer" de KNIME utilizando una matriz de confusión. Para el algoritmo MLP, se establecen los siguientes parámetros: (i) número máximo de iteraciones: 1000; (ii) número de capas ocultas: 2; y (iii) número de neuronas ocultas por capa: 10. En el caso del algoritmo PNN, se utilizan 2000 épocas. Para los demás algoritmos, se emplea la configuración predeterminada de cada módulo.

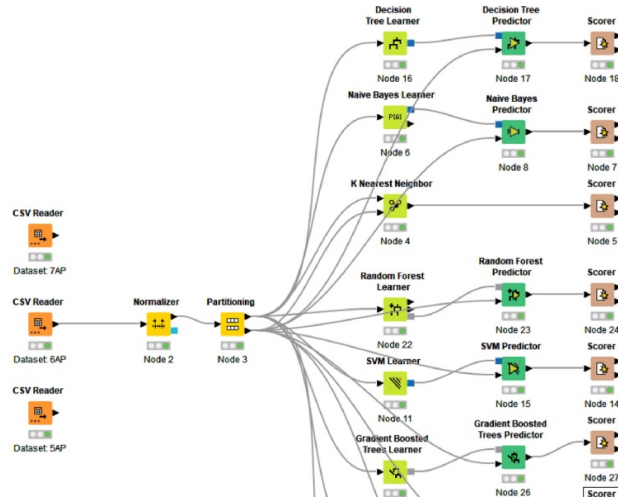


Figura 5: Flujo de trabajo implementado en KNIME con ocho algoritmos de machine learning.

5. Resultados Experimentales

En un análisis realizado en KNIME, se evaluaron varios algoritmos para resolver un problema de clasificación de habitaciones de terminales móviles. Los resultados destacan que el algoritmo Naive Bayes obtuvo la mejor precisión, con un impresionante 99,00 % utilizando 7 atributos predictivos y un 98,167 % con 5 atributos.

En segundo lugar se encuentra el algoritmo K-Nearest Neighbors (k-NN). Por otro lado, los algoritmos de redes neuronales artificiales, como MLP y PNN, mostraron una precisión inferior, con el MLP alcanzando un 96,833 % y el PNN solo llegando al 94,667 % (con 7 atributos predictivos).

Estos resultados se presentan en la Figura 6 como parte de varias simulaciones. En resumen, Naive Bayes demostró ser el algoritmo más efectivo en este contexto, seguido por k-NN, mientras que los algoritmos de redes neuronales tuvieron un rendimiento comparativamente menor en términos de precisión.

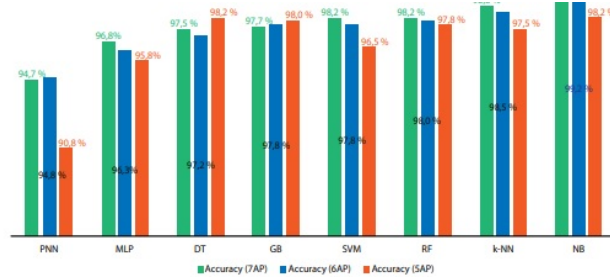


Figura 6: Resultados de la simulación realizada en KNIME.

6. Conclusiones

En esta investigación, se realizaron pruebas de clasificación de habitaciones de terminales móviles utilizando varios algoritmos. Los resultados destacan que el algoritmo Naive Bayes alcanzó la mayor precisión, con un impresionante 99 % de acierto, seguido por el algoritmo K-Nearest Neighbors (k-NN) con un 98.833 % de precisión cuando se usaron 7 atributos predictivos. Cuando se redujo el número de atributos, Naive Bayes mantuvo una alta precisión del 98.167 %.

En contraste, los algoritmos basados en Redes Neuronales Artificiales (ANN), como MLP y PNN, mostraron un rendimiento inferior, con MLP logrando un 96.833 % de precisión y PNN alcanzando solo el 94.667 % (con 7 atributos predictivos).

Se proponen futuras investigaciones, como la creación de datasets que incluyan mediciones de múltiples bandas de espectro para redes Wi-Fi y pruebas georreferenciadas para redes móviles con el fin de mejorar la precisión en la localización de dispositivos. Se enfatiza la importancia de una recolección adecuada de datos y se reconoce que los resultados pueden variar según el entorno interior específico. En resumen, esta investigación resalta la eficacia de Naive Bayes y k-NN en la clasificación de habitaciones de terminales móviles y plantea posibles direcciones para investigaciones futuras en el campo de la localización de dispositivos en entornos interiores.

Referencias

- [1] V. Bellavista-Parent, J. Torres-Sospedra, and A. Perez-Navarro. New trends in indoor positioning based on wifi and machine learning: A systematic review. In *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–8, 2021.
- [2] K. Sabanci, E. Yigit, D. Ustun, A. Toktas, and M. Aslan. Wifi based indoor localization: Application and comparison of machine learning algorithms. In *2018 XXIIIrd International Seminar/Workshop on Direct and Inverse*

- Problems of Electromagnetic and Acoustic Wave Theory (DIPED)*, pages 246–251, 2018.
- [3] N. Singh, S. Choe, and R. Punmiya. Machine learning based indoor localization using wi-fi rssi fingerprints: An overview. *IEEE Access*, 9:127150–127174, 2021.
 - [4] A. Tahat, R. Awwad, N. Baydoun, S. Al-Nabih, and T. A. Edwan. An empirical evaluation of machine learning algorithms for indoor localization using dual-band wifi. In *2021 2nd European Symposium on Software Engineering*, pages 106–111, 2021.
 - [5] J. Xue, J. Liu, M. Sheng, Y. Shi, and J. Li. A wifi fingerprint based high-adaptability indoor localization via machine learning. *China Communications*, 17(7):247–259, 2020.
 - [6] W. Xun, L. Sun, C. Han, Z. Lin, and J. Guo. Depthwise separable convolution based passive indoor localization using csi fingerprint. In *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2020.